Journal of
Computer & Electrical and Electronics Engineering Sciences
Original Article
DOI: 10.51271/JCEEES-0017

Steering angle prediction in autonomous vehicles: a deep learning approach combining VGG16 and LSTM

Ahmet Mehmet Karadeniz[1], Nuri Furkan Koçak[2]
[1]Doctoral School of Multidisciplinary Engineering Sciences, Széchenyi István University, Győr, Hungary
[2]Department of Electronics and Automation, Faculty of Engineering, Kırıkkale University , Kırıkkale, Turkiye

Cite this article: Karadeniz, A.M., & Koçak, N.F. (2024) Steering angle prediction in autonomous vehicles: a deep learning approach combining VGG16 and LSTM. *J Comp Electr Electron Eng Sci*, 2(2), 46-51.

Corresponding Author: Ahmet Mehmet Karadeniz, karadeniz.ahmet.mehmet@sze.hu

Received: 13/09/2024 ◆ **Accepted**: 27/09/2024 ◆ **Published**: 29/10/2024

ABSTRACT

The field of autonomous driving has seen remarkable progress in recent years, with the prediction of steering angles based on varying road conditions emerging as a critical area of focus. While previous efforts have concentrated on lane detection, road object identification, and 3-D reconstruction, our research centers on a vision-based model that leverages deep networks to translate raw camera images into steering angles without the need for predefined feature learning. In our paper, we introduce an end-to-end model that employs deep transfer learning to predict steering angles from image sequences captured by onboard cameras. This model merges two deep learning architectures: a convolutional neural network (CNN) and a long short-term memory (LSTM) network. We utilize the VGG16 network, pre-trained on ImageNet and renowned for its performance, to extract spatial features from the images. Concurrently, the LSTM network processes the temporal information embedded within the image sequences. Our proposed model comprehensively processes spatial-temporal data and adeptly models the nonlinear relationship between the input images and the steering angles. We conducted an experimental study using a publicly available dataset to evaluate the model's effectiveness. The outcomes of our experimental analysis reveal that our model delivers highly efficient and accurate steering angle predictions, effectively emulating human driving patterns. Moreover, our proposed model delivered highly efficient and accurate steering angle predictions, achieving a mean squared error (MSE) of 0.0728 on the validation dataset. This result outperforms conventional models, such as NVIDIA and 3D LSTM, in terms of both accuracy and training efficiency.

Keywords: Steering angle prediction, VGG16, LSTM, deep learning, convolutional neural networks

INTRODUCTION

The advent of autonomous vehicles represents one of the most transformative technological developments of the past decade, with far-reaching implications for road safety, efficiency, and transportation costs. Autonomous vehicles have the potential to significantly reduce traffic accidents, enhance fuel efficiency, and decrease the overall cost of vehicle ownership (Do et al., 2018; Badue et al., 2021). Recent advances in autonomous vehicle technology have been driven by rapid developments in computing power, image processing techniques, and sensor technologies (Velaskar et al., 2014).

Lane departure incidents are predominantly caused by driver distraction and constitute a substantial proportion of vehicular accidents. According to statistics, lane departures were responsible for 51% of traffic accidents in the United States in 2011 (Mammeri et al., 2015). To mitigate these risks, modern vehicles are increasingly being equipped with lane departure warning systems that rely on accurate steering angle estimation to ensure lane-keeping (Khodayari et al., 2010). The capacity to autonomously perform complex tasks-ranging from driving to surveillance and firefighting-underscores the significance of accurate steering angle prediction (Alshbatat, 2013; Saleem et al., 2021). Accurate steering angle estimation is a critical component of successful lane-keeping and overall vehicular control in these systems. It is recognized as a vital aspect of contemporary vehicle technologies, particularly in the context of developing autonomous driving systems and advanced driver assistance systems (ADAS). Accurate steering angle estimation provides real-time insights into a vehicle's steering dynamics, which is crucial for enhancing road safety and reducing the likelihood of accidents (Aparna et al., 2021).

Approaches to steering angle estimation generally fall into two categories: computer vision-based methods and learning-based methods (Oussama et al., 2020). Computer vision

This work is licensed under a Creative Commons Attribution 4.0 International License.

techniques involve the extraction of features from images to compute the steering angle, whereas learning-based methods utilize neural networks to predict the angle through end-to-end training (Gidado et al., 2020). End-to-end methods have the advantage of direct optimization without having to learn rules and to learn (Jiang et al., 2020). The integration of sophisticated machine learning algorithms, particularly convolutional neural networks (CNNs), has markedly enhanced the capabilities of autonomous driving systems (Song et al., 2022). CNNs have revolutionized pattern recognition by automating the feature extraction process from training data, thereby obviating the need for manually designed feature extraction stages. This automation has led to significant improvements in the accuracy of image-based applications, which are essential for tasks such as lane detection and steering angle prediction (Badue et al., 2021). However, most existing end-to-end steering angle prediction models use a single deep convolutional neural network from sensing to control. These methods can perform well in a single environment, but these models lack memory (Jiang et al., 2020).

In this paper, we propose an innovative steering angle prediction model that combines the Visual Geometry Group 16 (VGG16) model with the long short-term memory (LSTM) to address the drawbacks of existing models. Our model effectively utilizes spatial and temporal information. VGG16 is used to extract spatial features of input images, while LSTM is used to capture temporal dynamics. To reduce the training time, we utilize a transfer learning approach and use the VGG16 model pre-trained on the ImageNet dataset. This method provides faster and more efficient results than training the network from scratch.

Literature Review

Steering angle prediction is a key component in developing autonomous driving systems. VGG16 transfer learning has emerged as a promising approach to improve prediction accuracy while reducing computational complexity. VGG16, a convolutional neural network (CNN) architecture, is particularly effective in image classification tasks, making it suitable for processing visual data from cameras in autonomous vehicles (Song et al., 2022). Using transfer learning, VGG16 can use pre-trained models on large datasets, significantly improving training efficiency and accuracy in steering angle prediction tasks (Alsherif et al., 2023).

To further improve the robustness of the model, data augmentation techniques such as horizontal flipping and random angle adjustments are employed. These strategies diversify the training data set, helping to mitigate overfitting and improve performance in different driving scenarios, as demonstrated in road tests (Song et al., 2022). In addition, the integration of VGG16 with LSTM networks allows the model to capture temporal dependencies in sequential image data, leading to more accurate steering predictions. By integrating the dynamics of vehicle control over time, this approach improves steering angle prediction accuracy by approximately 95%, outperforming traditional CNN models (Hoang et al., 2023).

Despite the advantages of VGG16 and transfer learning, there are still challenges in real-time processing and adaptation to different driving environments. These limitations highlight the need for further research and development in this area to improve the practical application of these models in autonomous vehicles.

In particular, the transfer learning approach with VGG16 is particularly beneficial in scenarios where labeled data is limited. By fine-tuning VGG16 to specific driving datasets, researchers can achieve improved steering angle prediction as the model retains learned features from large image datasets. The depth and architecture of VGG16 allow it to capture intricate patterns in visual data, which is critical for understanding and predicting driving scenarios. In addition, transfer learning reduces the dependence on large datasets, making it possible to train models in environments where data collection is difficult (Ismail et al., 2024).

Studies have shown that models using VGG16 for steering angle prediction can achieve high accuracy, with performance metrics such as precision and recall significantly improved over traditional methods (Golnari et al., 2024; Karadeniz et al., 2024). However, it is important to explore alternative architectures such as DenseNet or ResNet, which may offer competitive performance and efficiency in certain contexts (Ismail et al., 2024).

In this study we have gone through the following questions and answered them in Section 3:

RQ1: How does combining VGG16 with LSTM improve steering angle prediction?

RQ2: What are the key factors that contribute to the accuracy and efficiency of the model?

RQ3: What are the potential limitations of the model in different driving environments?

METHODS

Convolutional Neural Networks

Over recent years, it has become clear that convolutional neural networks (CNNs) are highly effective at generating a detailed representation of an input image by transforming it into a fixed-length vector. This representation is versatile and can be applied to a range of visual tasks. CNNs excel at processing visual imagery. Typically, a CNN is composed of an input layer, an output layer, and several hidden layers. The hidden layers include a sequence of convolutional layers, pooling layers, normalization layers, and fully connected layers. The convolutional layers perform operations on the data (often images) using multiplication or a dot product. Although these operations are traditionally called convolutions, they are mathematically carried out by a sliding window or filter executing a dot product with the image. This process is crucial for determining how weights are assigned at specific index points within the matrix (Turk, 2024). The resulting output from the convolutional layer is passed on to the subsequent pooling layer, whose role is to downsize the data to simplify the computation. Owing to their exceptional ability to interpret visual data, CNNs and their variations are now being utilized for predicting steering angles. Researchers have investigated various CNN architectures, which differ in the number and size of layers and neurons, to optimize steering angle prediction (Song et al., 2022; Turk, 2024).

LSTMs, or long short-term memory networks, differ from traditional feedforward neural networks by having feedback connections. This allows LSTMs to not only process

individual data points, like images, but also to handle entire sequences of data, such as speech or video streams. LSTMs have demonstrated superior capabilities in learning and understanding sequences, which makes them suitable for autonomous vehicles (AVs) to maintain lane positions by analyzing the relationship between consecutive image frames. Typically, LSTMs are utilized after a CNN, which is responsible for feature detection, to comprehend temporal relationships within the data. This combined CNN-LSTM model has been adopted in various research studies (Saleem et al., 2021).

Data Processing

The dataset used in this study can be found in Kaggle as shown in. It includes approximately 10000 images along with their associated steering angle data. Figure 1 showcases a histogram of the collected data, depicting the distribution of steering angles against the frequency of their occurrence in each bin.

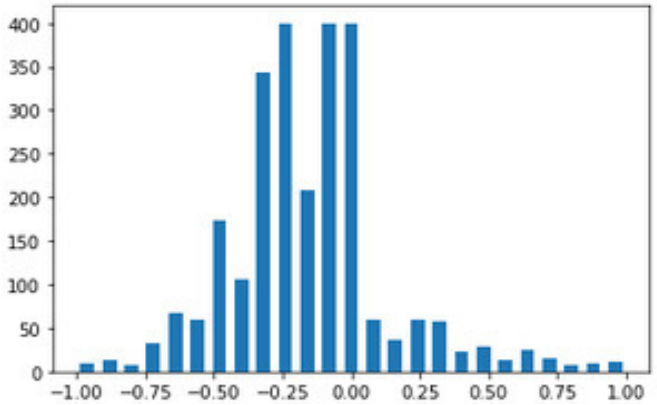

Figure 1. Steering angles and their frequencies

After collecting the driving data, it undergoes several preprocessing steps to make it suitable for training purposes. These steps involve a range of techniques such as data cleansing, image cropping and resizing as well as removing superfluous data and dealing with any missing values. The camera images are cropped to focus on the road lane markings which are the area of interest. Then, these cropped images are resized to conform to the specific dimensions required for the model's input. Preprocessing the images facilitates faster processing and ensures the model focuses on extracting only the relevant datasets. Figure 2, 3 illustrate the images from the front-facing camera before the preprocessing steps and after, respectively.

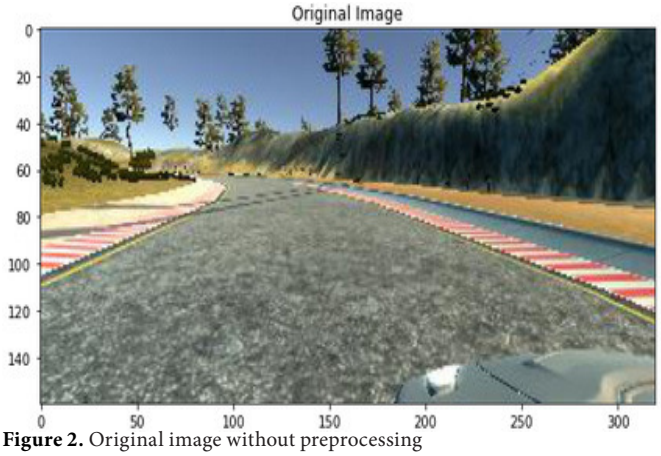

Figure 2. Original image without preprocessing

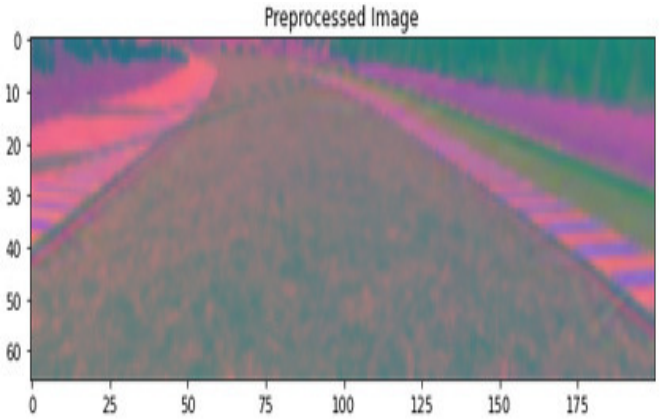

Figure 3. Preprocessed image

After preprocessing, the data is divided into two distinct subsets: 80% of it for the training and 20% for the validation. The training dataset is used to train the deep learning model and the validation dataset allows for performance evaluation during the training process and also it is used to evaluate the model's final performance. This split is a standard deep learning technique for preventing overfitting and underfitting. The Sci-kitlearn library has been utilized to randomly partition the data using Pandas and NumPy tools. Figure 4 illustrates the common steering angles across the training and validation datasets.

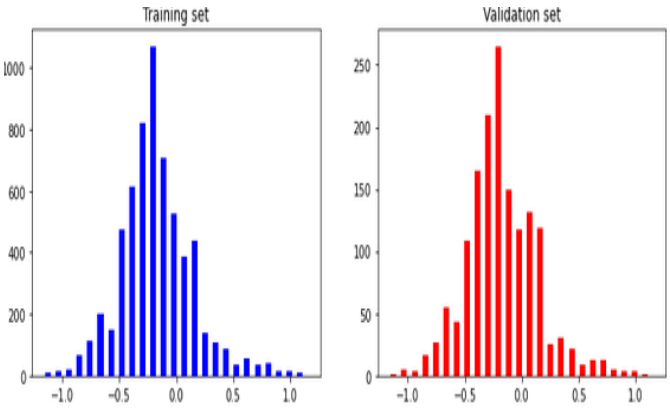

Figure 4. Training and validation datasets

Model Architecture

The next phase involves defining the model architecture. For this purpose, the VGG16 transfer learning model, which has been pre-trained on the ImageNet dataset and is available in the Tensorflow. Keras library, is chosen. The base layers of the VGG16 are frozen to inhibit any additional training. To customize the model for the steering angle prediction task, custom layers are integrated on top of the VGG16 base. An LSTM layer is added to process sequential data and capture temporal dependencies, which is crucial for understanding sequences like video frames. This is followed by several dense layers, including fully connected layers with 100 neurons, 50 neurons, 10 neurons, and finally 1 neuron to output the predicted steering angle. These layers employ as the activation function an exponential linear unit (ELU) to introduce nonlinearity. The culmination of the model architecture is a solitary neuron that serves as the output layer, tasked with predicting the steering angle. The structural designs of the model, including the integration of VGG16 and LSTM layers as well as number of parameters, are depicted in Figure 5.

Layer (type)	Output Shape	Param #
time_distributed_8 (TimeDist	(None, 5, 2, 6, 512)	14714688
time_distributed_9 (TimeDist	(None, 5, 1, 3, 24)	307224
time_distributed_10 (TimeDis	(None, 5, 1, 2, 36)	21636
time_distributed_11 (TimeDis	(None, 5, 1, 1, 48)	43248
time_distributed_12 (TimeDis	(None, 5, 1, 1, 64)	27712
time_distributed_13 (TimeDis	(None, 5, 1, 1, 64)	36928
time_distributed_14 (TimeDis	(None, 5, 1, 1, 64)	0
time_distributed_15 (TimeDis	(None, 5, 64)	0
lstm_2 (LSTM)	(None, 5, 100)	66000
dropout_5 (Dropout)	(None, 5, 100)	0
lstm_3 (LSTM)	(None, 50)	30200
dropout_6 (Dropout)	(None, 50)	0
dense_2 (Dense)	(None, 10)	510
dropout_7 (Dropout)	(None, 10)	0
dense_3 (Dense)	(None, 1)	11

Total params: 15,248,157
Trainable params: 533,469
Non-trainable params: 14,714,688

Figure 5. Detailed architecture of the proposed steering angle prediction model, combining VGG16 for spatial feature extraction and LSTM for temporal dependency handling

VGG16: Visual Geometry Group 16, LSTM: Long short-term memory

Evaluation Metrics and Training Process

During the training phase, the preprocessed data is input into the model for learning. The TensorFlow and Keras frameworks facilitate the training of the models. Throughout this phase, hyperparameters such as batch size and learning rate are fine-tuned. As a loss function mean squared error (MSE) is chosen. Then the Adam optimizer is leveraged to reduce the MSE loss, which quantifies the difference between the predicted steering angles and the actual values. Key performance indicators, specifically training and validation loss are monitored to gauge the model's effectiveness during the training. Upon completion of the training, the model is saved for future use in determining the steering angles for the autonomous vehicle. To capture the best model weights, a checkpoint callback function is utilized which ensures the model is saved at the point of lowest loss throughout the epochs.

The effectiveness of the models is evaluated using the test dataset. The MSE is used as the metric for evaluation. The optimal value for MSE is zero, indicating perfect predictions with no errors. The mathematical expression for calculating MSE is presented in equation (1).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (1)$$

Where:

n is the number of samples,

ŷi represents the predicted steering angle,

yi represents the actual steering angle,

Σ indicates summation over all samples.

The mean squared error (MSE) is selected as the primary evaluation metric due to its sensitivity to large errors, which is crucial in steering angle prediction. Large deviations in steering angles can lead to lane departure or accidents, making MSE an effective measure of model performance. Furthermore, MSE penalizes larger errors more heavily, ensuring that the model focuses on minimizing significant prediction deviations. The difference between training and validation losses indicates the model's generalization ability. In our experiments, the small gap between these losses demonstrates that the model did not overfit to the training data, validating its robustness across different datasets.

RESULTS AND DISCUSSION

In our study, we utilized a Kaggle notebook environment consisting of GPU P100 accelerator and Keras for training and validating our neural networks. We used a dataset from Kaggle and provided it in, comprising 1802 training samples and 451 validation samples due to limited computational resources. During the training phase, the model underwent 20 epochs with a learning rate set at 0.001 with Adam optimizer and a batch size of 32. The losses of MSE score per epochs are demonstrated in Figure 6. Remarkably, the training was completed in just 41 minutes, demonstrating a more efficient use of time compared to other models.

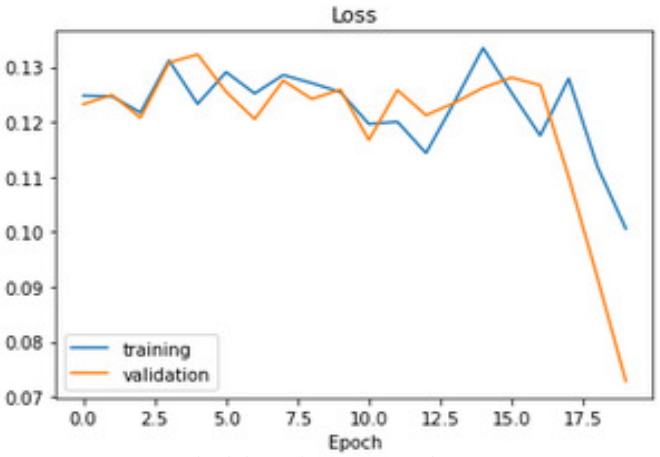

Figure 6. Training and validation losses per epochs

The robustness of our model is further enhanced through data augmentation techniques such as horizontal flipping and random angle adjustments. These methods ensure the model can generalize well across different driving scenarios, reducing the risk of overfitting. Furthermore, the integration of temporal dependencies via LSTM networks allows the model to handle sequential data more effectively, improving performance under real-world conditions.

Limitations

Also, the proposed model combining VGG16 and LSTM resulted 0.0728 MSE score in the validation dataset and its comparison with other relevant studies are given in Table. According to the comparison, our model surpassed the NVIDIA and 3D LSTM, however performed slightly worse against VGG16. The reason of it might be because of the used dataset size, computational resources such as accelerators (GPU) or the limitation of the dataset used. For example, if a dataset is chosen which includes not diverse values, then obviously the loss will be much lower.

Table. Comparison of the models

Model	MSE score
NVIDIA (Du et al., 2019)	0.3154
3D LSTM (Du et al., 2019)	0.3374
VGG16 (Jiang et al., 2020)	0.0413
Our proposed model	0.0728

MSE: Mean squared error, VGG16: Visual Geometry Group 16, LSTM: Long short-term memory

CONCLUSION

In this study, we introduce an End-to-End steering angle prediction model that fuses a modified VGG16 network with LSTM to effectively harness both spatial and temporal information from input image sequences. The modified VGG16, based on transfer learning, allows for efficient training, while the LSTM network captures crucial temporal data. Our mixed network design, validated on the publicly available dataset demonstrates accurate steering angle predictions and robustness across different environments, as evidenced by our model's lower loss value in terms of MSE score for the transfer learning and 3D convolutional models.

The visualization of the convolutional layers reveals that our model can learn key road features without predefined parameters an advantage of the End-to-End approach. Despite the promising results, these models have not yet been tested in real-world conditions or specialized datasets, an endeavor we aim to undertake in future work. We believe the robustness shown in experimental results bodes well for real-time autonomous driving applications.

Looking ahead, we see potential in expanding the deep learning model's architecture with larger and deeper layers, which may yield improved outcomes despite current computational limitations. However, there remains significant research to be done before such models can be widely deployed. Future models could benefit from training on more diverse data. Additionally, a high-quality simulator coupled with deep reinforcement learning, guided by a reward function prioritizing efficiency, ride smoothness, rule adherence, and safety, could further refine these models for public transportation use.

Benefits and Drawbacks

Every approach has advantages and disadvantages. We also went through analysis of the benefits and drawbacks of our approach and they can be found as the following:

Benefits: The combination of VGG16 and LSTM captures both spatial and temporal features, leading to improved steering angle prediction accuracy.

Transfer learning with VGG16 significantly reduces training time.

The model outperforms other architectures like NVIDIA and 3D LSTM in certain scenarios.

Drawbacks: The model has not been tested in real-world driving environments, and the dataset used may not represent all possible driving conditions.

Computational resource limitations restricted the dataset size and could have impacted model performance.

ETHICAL DECLARATIONS

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Do, T.D., Duong, M.T., Dang, Q.V., & Le, M.H. (2018). Real-time self-driving car navigation using deep neural network. In 2018 4th International Conference on Green Technology and Sustainable Development (GTSD) (pp. 7-12). IEEE. doi:10.1109/GTSD.2018.8595590
2. Badue, C., Guidolini, R., Carneiro, R.V., Azevedo, P., Cardoso, V.B., Forechi, A., ... & De Souza, A.F. (2021). Self-driving cars: a survey. *Expert Systems Appl,* 165,113816. doi:10.1016/j.eswa.2020.113816
3. Velaskar, P., Vargas-Clara, A., Jameel, O., & Redkar, S. (2014). Guided navigation control of an unmanned ground vehicle using global positioning systems and inertial navigation systems. *Int J Electr Comput Engineer,* 4(3),2088-8708. doi:10.11591/ijece.v4i3.5183
4. Mammeri, A., Lu, G., & Boukerche, A. (2015). Design of lane keeping assist system for autonomous vehicles. In 2015 7th International Conference on New Technologies, Mobility and Security (NTMS) (pp. 1-5). IEEE. doi:10.1109/NTMS.2015.7266483
5. Khodayari, A., Ghaffari, A., Ameli, S., & Flahatgar, J. (2010). A historical review on lateral and longitudinal control of autonomous vehicle motions. In 2010 International Conference on Mechanical and Electrical Technology (pp. 421-429). IEEE. doi:10.1109/ICMET.2010.5598396
6. Alshbatat, A.I.N. (2013). Behavior-based approach for the detection of land mines utilizing off the shelf low cost autonomous robot. *Int J Robotics Automation,* 2(3),83. doi:10.11591/ijra.v2i3.2038
7. Saleem, H., Riaz, F., Mostarda, L., Niazi, M.A., Rafiq, A., & Saeed, S. (2021). Steering angle prediction techniques for autonomous ground vehicles: a review. *IEEE Access,* 9,78567-78585. doi:10.1109/ACCESS.2021.3083890
8. Aparna, M.P., Gandhiraj, R., & Panda, M. (2021). Steering angle prediction for autonomous driving using federated learning: The impact of vehicle-to-everything communication. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE. doi:10.1109/ICCCNT51525.2021.9580097
9. Oussama, A., & Mohamed, T. (2020). A literature review of steering angle prediction algorithms for self-driving cars. *Advan Intell Syst Appl Comput Sci,* 4,30-38. doi:10.1007/978-3-030-36674-2_4
10. Gidado, U.M., Chiroma, H., Aljojo, N., Abubakar, S., Popoola, S.I., & Al-Garadi, M.A. (2020). A survey on deep learning for steering angle prediction in autonomous vehicles. *IEEE Access,* 8,163797-163817. doi:10.1109/ACCESS.2020.3017883
11. Jiang, H., Chang, L., Li, Q., & Chen, D. (2020). Deep transfer learning enable end-to-end steering angles prediction for self-driving car. In 2020 IEEE Intelligent Vehicles Symposium (IV) (pp. 405-412). IEEE. doi:10.1109/IV47402.2020.9304611
12. Song, X., Cao, H., & You, H. (2022, December). Steering Wheel Rotation Angle Prediction Based on VGG-16 and Data Augmentation. In 2022 4th International Conference on Control and Robotics (ICCR) (pp. 398-402). IEEE. doi:10.1109/ICCR55715.2022.10053867
13. Alsherif, M., Daowd, M., Bassiuny, A.M., & Metered, H.A. (2023). Utilizing transfer learning in the udacity simulator to train a self-driving car for steering angle prediction. In 2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS) (pp. 134-139). IEEE. doi:10.1109/ICICIS58388.2023.10391185

14. Song, X., Cao, H., & You, H. (2022). Steering wheel rotation angle prediction based on VGG-16 and data augmentation. In 2022 4th International Conference on Control and Robotics (ICCR) (pp. 398-402). IEEE. doi:10.1109/ICCR55715.2022.10053867
15. Hoang, T.N., Hong, P.P., Vinh, N.N., Nguyen, N.T., Nguyen, K.H., & Quach, L.D. (2023). An improved lane-keeping controller for autonomous vehicles leveraging an integrated CNN-LSTM approach. *Int J Advan Comput Sci Appl,* 14,7. doi:10.14569/IJACSA.2023.0140723
16. Ismail, S.N.M.S., Razak, S.F.A., & Aziz, N.A.A. (2024). Transfer learning for improved electrocardiogram diagnosis of cardiac disease: exploring the potential of pre-trained models. *Bullet Electr Engineer Informat,* 13(5),3288-3300. doi:10.11591/eei.v13i5.7053
17. Ismail, S.N.M.S., Razak, S.F.A., & Aziz, N.A.A. (2024). Transfer learning for improved electrocardiogram diagnosis of cardiac disease: exploring the potential of pre-trained models. *Bullet Electr Engineer Informat,* 13(5),3288-3300. doi:10.11591/eei.v13i5.7053
18. Golnari, A., Komeili, M.H., & Azizi, Z. (2024). Probabilistic deep learning and transfer learning for robust cryptocurrency price prediction. *Expert Syst Appl,* 124404.doi:10.1016/j.eswa.2024.124404
19. Karadeniz, A. M., Ballagi, Á., & Kóczy, L. T. (2024). Transfer learning-based steering angle prediction and control with fuzzy signatures-enhanced fuzzy systems for autonomous vehicles. *Symmetry,* 16(9),1180. doi:10.3390/sym16091180
20. Turk, F. (2024). Investigation of machine learning algorithms on heart disease through dominant feature detection and feature selection. *SIViP,* 18,3943-3955. doi:10.1007/s11760-024-03060-0
21. Turk F. (2024). RNGU-NET: a novel efficient approach in segmenting tuberculosis using chest X-ray images. *PeerJ Comput Sci,* 10:e1780 doi:10.7717/peerj-cs.1780
22. https://www.kaggle.com/datasets/zaynena/selfdriving-car-simulator (accessed on 25 August 2024).
23. Du, S., Guo, H., & Simpson, A. (2019). Self-driving car steering angle prediction based on image recognition. arXiv preprint arXiv:1912.05440. doi:10.48550/arXiv.1912.05440